A Simple Technique for Self-Calibration

Paulo R. S. Mendonça and Roberto Cipolla
Department of Engineering
University of Cambridge
Trumpington Street, Cambridge, UK, CB2 1PZ
{prdsm2,cipolla}@eng.cam.ac.uk

Abstract

This paper introduces an extension of Hartley's self-calibration technique [8] based on properties of the essential matrix, allowing for the stable computation of varying focal lengths and principal point. It is well known that the three singular values of an essential must satisfy two conditions: one of them must be zero and the other two must be identical. An essential matrix is obtained from the fundamental matrix by a transformation involving the intrinsic parameters of the pair of cameras associated with the two views. Thus, constraints on the essential matrix can be translated into constraints on the intrinsic parameters of the pair of cameras. This allows for a search in the space of intrinsic parameters of the cameras in order to minimize a cost function related to the constraints. This approach is shown to be simpler than other methods, with comparable accuracy in the results. Another advantage of the technique is that it does not require as input a consistent set of weakly calibrated camera matrices (as defined by Hartley) for the whole image sequence, i.e., a set of cameras consistent with the correspondences and known up to a projective transformation.

1. Introduction

The problem of self-calibration has attracted the attention of researchers in the computer vision community as a powerful method that allows for the recovery of 3D models from image sequences. Compared to the classical calibration problem [21, 22, 3], the algorithms for self-calibration make no or few assumptions about the particular structure of the scene being viewed. Instead, they attempt to find the intrinsic parameters of the cameras exploiting constraints imposed over these parameters from epipolar or trilinear relations or, from a set of camera matrices, run a numerical minimization on the space of 3D projective matrices that transforms the original set of cameras into a new one where the constraints on the intrinsic parameters are satisfied.

In [14] the epipolar constraints are imposed over the image of the absolute conic, which encodes the intrinsic parameters. When the camera motion is restricted to be planar – which, incidentally, is a critical motion for self-calibration [17, 23] – image triplets have to be used, as presented in [1], where the intrinsic parameters are found from constraints arising from properties of the trifocal tensor [10, 11, 19]. A closed form solution for the horizontal and vertical scale factors was found in [2], which is then refined by a search in the space of projective transformations. This is similar to what is presented in [15], where properties of the absolute quadric [20] are exploited. A common characteristic to the self-calibration techniques based in seaching for an appropriate 3D projective transformation is the use of a set of weakly calibrated projective cameras [9]. This is defined as a set of cameras which is consistent with the image to image correspondences and is defined up to a 3D projective transformation. For image pairs and triplets, these can be found from the fundamental matrix or the trifocal tensor. The authors are not aware of any technique to directly obtain a set of consistent cameras from a quadfocal tensor [4]. When more then four images are available bundle adjustment [18] becomes necessary. This is a computationally expensive technique, as it minimizes the geometric error of the reprojected points as a function of the entries of the camera matrices and the coordinates of the points in space.

2. Self-Calibration from the Essential Matrix

A novel approach was introduced by Hartley in [8]. In [12] it is proved that two of the three singular values of the essential matrix must be equal. As the essential matrix has rank two, the remaining singular value must be zero. These are necessary and sufficient conditions for the decomposition of the essential matrix in the rotation and the direction of the translation relating the associated pair of cameras. In [12] it is also shown that the equal singular value condition imposes not one, but two algebraic constraints over the entries of the essential matrix. This fact was exploited by Hartley for self-calibration, since the essential matrix $\mathbf{E}$ depends on the intrinsic parameters of the cameras, encoded in the calibration matrices $\mathbf{K}_1$ and $\mathbf{K}_2$, as well as on the fundamental matrix $\mathbf{F}$, according to

$$\mathbf{E} = \mathbf{K}_2^{\mathrm{T}} \mathbf{F} \mathbf{K}_1. \tag{1}$$

Since the fundamental matrix has rank two and the calibration matrices have full rank, the condition that the essential matrix has rank two is automatically satisfied. Nevertheless, the equal singular value condition is not satisfied when the matrices $\mathbf{K}_1$ and $\mathbf{K}_2$ are arbitrary, and thus two parameters can be computed. It should be clear now that the equal singular value condition establishes a link between the motion and the intrinsic parameters of the associated pair of cameras, in the same sense that the Kruppa's equations do [14]. Assuming that the principal points are at the centre of the images, the skew is zero and the aspect ratio of the cameras is known, it is possible to compute the scale factors (focal length times magnification) of both cameras and then, from the essential matrix, recover the motion parameters.

A drawback of this algorithm is that one can only obtain two of the camera parameters, and its robustness and stability are inferior to the ones found in techniques based on numerical optimization of the entries of the 3D projective transformation matrices. On the other hand, it is simpler to implement and to analyse, as there is no reference to conics or quadrics with imaginary axes – see [15, 20]!

The novel idea introduced in this paper is to extend Hartley's results for larger image sequences, what, as will be shown, allows for the computation of more and varying intrinsic parameters. This goal is achieved by solving an optimization problem by numerical techniques, searching directly for the intrinsic parameters of the cameras, instead of the indirect search performed by the algorithms based on the projective transformation. This leads to a simple, practical and more reliable approach, that makes use of all information available. The technique does not require as input a set of weakly calibrated camera matrices, and since it merges information from all available image pairs, the consistency of the cameras is embedded in the algorithm.

3. Description of the Algorithm

Let n_k be the number of known intrinsic parameters of each camera, and n_f the number of unknown but fixed intrinsic parameters (typically, the aspect ratio) of the cameras. As pointed in [16], the number of images that allow for self-calibration in this situation is n, where

$$n \times n_k + (n-1) \times n_f \geq 8. \tag{2}$$

This equation comes from the fact that each known intrinsic parameter introduces $n \times n_k$ constraints over the projective transformation that has to be applied to the weakly calibrated set of cameras to calibrate them, while the fixed parameters impose only $(n-1) \times n_f$ constraints. Any self-calibration scheme calibrates the cameras up to a 3D similarity transformation, that has 7 degrees of freedom (d.o.f.) (6 for the 3D Euclidean transformation plus 1 for the arbitrary scale). Since the projective space has 15 d.o.f., there are still 8 d.o.f. to be imposed, resulting in (2).

Assuming now that a sequence with n images is acquired. By matching points pairwise it is possible to find $n(n-1)/2$ fundamental matrices. Although the fundamental matrices are not independent [5], numerical stability and robustness are greatly improved when redundant data is used. Let $\mathbf{F}_{ij}$ be the fundamental matrix relating images i and j of the sequence, and let $\mathbf{K}_i$ and $\mathbf{K}_j$ be the calibration matrices of cameras i and j, parametrized as

$$\mathbf{K} = \begin{bmatrix} \alpha_x & s & u_0 \\ 0 & \varepsilon\alpha_x & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{3}$$

where α_x is the product of focal length and magnification factor, ε is the aspect ratio, $[u_0 \quad v_0]^{\mathrm{T}}$ are the coordinates of the principal point and s is the skew.

Let ${}^1\sigma_{ij}$ and ${}^2\sigma_{ij}$ be the non zero singular values of $\mathbf{K}_j^{\mathrm{T}} \mathbf{F}_{ij} \mathbf{K}_j$, in descending order. It is possible to establish a cost function C to be minimized in the entries of $\mathbf{K}_i, i = 1, ..., n$ as

$$C(\mathbf{K}_i, i = 1, ..., n) = \sum_{ij}^{n} \frac{w_{ij}}{\sum_{kl}^{n} w_{kl}} \frac{{}^1\sigma_{ij} - {}^2\sigma_{ij}}{{}^2\sigma_{ij}}, \tag{4}$$

where w_{ij} is a degree of confidence in the estimation of the fundamental matrix $\mathbf{F}_{ij}$, for example the inverse of the mean geometric distance between the image points and their corresponding epipolar lines. Other possibility is to make the weights w_{ij} equal to the number of points used in the computation of the fundamental matrix $\mathbf{F}_{ij}$. The derivatives of (4) can be computed accurately by finite differences, since the function that relates the entries of a matrix with its singular values is notably smooth. In fact, the Wielandt-Hoffman theorem for singular values [6] states that, if $\mathbf{A}$ and $\mathbf{E}$ are matrices in $\mathbb{R}^{m \times n}$ with $m \geq n$, then

$$\sum_{k=1}^{n} (\sigma_k(\mathbf{A} + \mathbf{E}) - \sigma_k(\mathbf{A}))^2 \leq \|\mathbf{E}\|_F^2, \tag{5}$$

where $\sigma_i(\mathbf{M})$ denotes the ith largest singular value of $\mathbf{M}$ and $\|\mathbf{M}\|_F$ is the Frobenius norm of $\mathbf{M}$. It means that if one perturbs a matrix $\mathbf{A}$ by adding to it an error matrix $\mathbf{E}$, the correspondent perturbation in any singular value of $\mathbf{A}$ will be smaller than the magnitude of $\mathbf{E}$ under the Frobenius norm.

The redundancy present in the minimization problem not only reinforces the numerical stability and robustness, but avoids bias towards any given image, which could have more influence in the cost function if more fundamental matrices built from it were used. With the presented formulation the cost function is completely symmetric in terms of the image points, due to the weighting factors, and the cameras. It is worth noting that this symmetry is hardly achieved by any other algorithm, except when bundle adjustment is used.

4. Practical Aspects of Self Calibration

It has become well established that self-calibration of the principal point is not only difficult, but also of low importance: the 3D reconstruction error after triangulation is not dramatically affected by errors in the assumption that the principal point remains fixed at the centre of the images. It is also not necessary to introduce more complications to the model by assuming that the skew is not zero. Calibration algorithms using calibration grids often attempt to ensure that $s = 0$ in (3), sometimes at the cost of solving a nonlinear optimization problem [3]. The aspect ratio, on the other hand, is not always 1 as is commonly assumed. Nevertheless, whatever is the true value of the aspect ratio, this is a remarkably stable parameter, and will not change due to zooming or focusing. If the cameras used in the acquisition of the images are known or are available, it is possible to calibrate for this parameter once, and reliably use the value found in any other occasion. For images of unknown origin, this procedure is not possible, and the assumption of the aspect ratio being equal to 1 cannot be blindly used. In the experiments that follow the skew will always assumed to be zero. The aspect ratio was found to be very close to 1 in previous calibration experiments, and thus this value was also used in all experiments with real data. In the experiment with the calibration grid the principal point was assumed to be in the centre of the image. In the experiment with the sequence of images of the building there are sufficient images, according to (2), to allow for the principal point to vary from image to image, as was actually done.

5. Experimental Results

5.1. Experiments with Synthetic Data

Firstly, the robustness of the proposed method was tested with synthetic data. Noise was added to the coordinates of the images of a set points, affecting the computation of the fundamental matrices related to the images, which were computed with a simple linear algorithm [11]. The estimative of the aspect ratio turned out to be very robust, with errors around 5% even for noise levels of 10 pixels to both coordinates of the points used to compute the fundamental matrix. The focal length is more sensitive to noise, producing errors around 20% for this same noise level. It is important to notice, however, that this does not mean that the reconstruction obtained from these parameters will be similarly affected. In fact, it is shown in [2] that these differences in the internal parameters can be explained in terms of the position of the plane at infinity. Essentially, if an affine approximation is valid, the focal length becomes meaningless. Ratios of focal lengths will still be important, but it was observed in the experiments that the ratio of focal lengths behaves exactly as the aspect ratio, being considerably insensitive to noise.

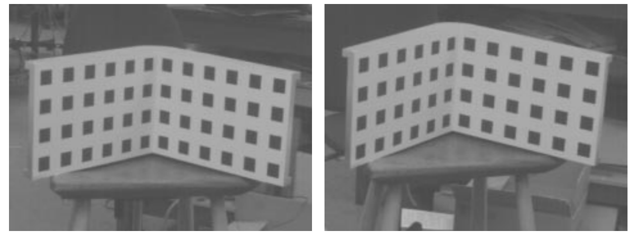

Figure 1. Calibration grids used in the self calibration algorithm for comparison with ground truth. The corners and edges of the black squares were automatically detected and matched between the two images. Note the small baseline and the absence of noticeable perspective effects, making the self-calibration problem very challenging.

5.2. Experiments with Real Data

A first experiment to show the quality of the reconstruction obtained with the parameters estimated by the technique presented here was carried out with the images shown in fig. 1. Since only two images were used, the algorithm falls in the case presented by Hartley, but instead of his closed form solution the cost function (4) was minimized using the Davidon-Fletcher-Powell (DFP) method [13]. The assumptions were that the skews were zero, the aspect ratios were one, and the principal points were at the centre of the images. Corners where found by using the Harris corner detector [7] and matched by cross-correlation, and the fundamental matrix relating the views was estimated by a linear technique [11]. Table 1 shows a comparison between the intrinsic parameters found for the first camera by using the the metric information provided by the calibration grid and the linear algorithm described in [3], and the intrinsic parameters found by the self-calibration algorithm introduced here. See image caption for notation.

	Calibration using the grid	Self-calibration	Relative error
α_x	1945.04	2137.72	9.91%
ε	0.98	1.00	1.72%
θ	89.98°	90.00°	0.02%
u_0	349.59	320.00	-8.46%
v_0	187.80	240.00	27.80%

Table 1. Comparison of the results of calibration using metric information provided by the calibration grid and the algorithm presented here, both applied to the images shown in fig. 1. The notation is the same as in (3), except that the angle θ between the axis was computed from the skew factor, as described in [3].

After self-calibrating the cameras, reconstruction up to a similarity transformation was possible by a triangulation technique. The reconstructed squares are shown in fig. 2.

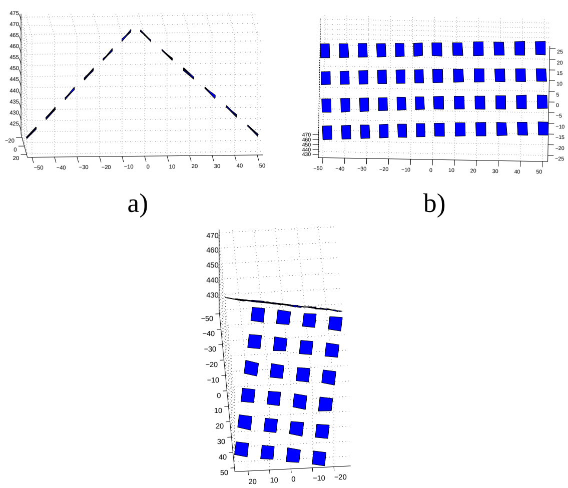
a) b) c)

Figure 2. Top, front and side views of the reconstructed grid.

As a measure of the accuracy of the reconstruction, the angle between the planes of the reconstructed grid was computed, resulting in 87° degrees, very close to the 90° of the real grid. As can be seen from fig. 1, the baseline is very small, and the perspective effects are very weak. As pointed in [2], these conditions make the self-calibration problem very difficult. Even though, satisfactory results found.

The algorithm was also tested in the sequence of images shown in fig. 3. The reconstruction of a few points in the

Figure 3. Building sequence. Corners were detected automatically and matched by hand. The reconstruction of some points is shown in fig. 4.

Image	α_x	u_0	v_0
1	630.5	346.1	270.6
2	639.7	325.7	241.9
3	632.9	333.7	257.3
4	547.4	370.5	210.5
5	561.2	275.6	252.3
6	604.0	331.3	235.6

Table 2. Intrinsic parameters found from the images shown in fig. 3. The variation of the focal length is compatible with the zooming and focusing suggested by the motion, in particular the zooming out of the fourth view. The principal point varies quite unpredictably, but this behaviour does not affect the reconstruction.

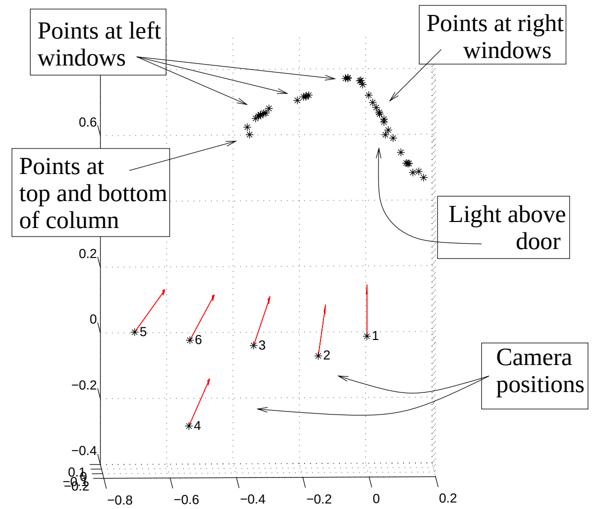

Figure 4. Top view of some points reconstructed from the building sequence. The appearance of the reconstruction is very good. Even the top of the light above the door on the right was correctly reconstructed. The column which contains the points indicated on the right becomes thicker from the top to the bottom, what was also recovered by the algorithm.

main left and right planes can be seen from a top view in fig. 4. In this experiment both the scale factor and the principal point were allowed to vary from image to image. The sequence of values assumed by these parameters is shown in tab. 2. The variation of the scale factor is in accordance with what is expected from the zooming and focusing required to keep a sharp image from the different positions from where the snapshots were taken. The principal point varies considerably around the centre of the image, in agreement with the results presented in [2]. In fig. 5 it is shown the result of reprojecting the reconstructed points back to the first image of the sequence. The mean geometric error found in the computation of the fundamental matrix directly from the correspondences was 0.77 pixels. The direct computation of the fundamental matrix from the calibrated camera matrices reduced this error to 0.68. This improvement is a direct result of the merging of information among the images, and could not be obtained by any technique based simply in a search for a 3D projective transformation to be applied to a set of weakly calibrated cameras. In this case, since the fundamental matrix is invariant to any common 3D projective transformation applied to the associated camera pair, the geometric error computed above would have remained constant.

a) b)

Figure 5. Reprojection of the reconstructed points. a) The original mean geometric error (distance from cornerss to epipolar lines) computed from the fundamental matrices used in the reconstruction is 0.7682. b) The geometric error for the projection of the reconstructed points is 0.6837, showing the improvement resultant from the integration of the data among the images.

6 Conclusions

This article presented a new self-calibration technique extending the method developed by Hartley in [8] for the self-calibration of the focal lengths of a pair of cameras to the case of multiple varying intrinsic parameters and larger image sequences. The formulation of the algorithm is simple, corresponding to a direct search in the space of intrinsic parameters aiming to minimize a cost function based in properties of essential matrices. This method contrasts with the approach of searching for the projective transformation that maps the weakly calibrated camera matrices into a new set of camera matrices where given constraints over the intrinsic parameters are satisfied. Experiments with both synthetic and real data showed that the technique developed is numerically stable, robust and accurate. It dispenses the use of bundle adjustment to search for a consistent set of weakly calibrated camera matrices, has no bias towards any particular image of the sequence, and makes use of the information provided by the correspondences in an homogeneous way.

References

[1] M. Armstrong, A. Zisserman, and R. Hartley. Self-calibration from image triplets. In *Proc. 4th European Conf. on Computer Vision*, Lecture Notes in Computer Science 1064, pages 3–16, 1996.

[2] S. Bougnoux. From projective to euclidean space under any practical situation, a criticism of self-calibration. In *Proc. 6th Int. Conf. on Computer Vision*, pages 790–796, 1998.

[3] O. Faugeras. *Three-Dimensional Computer Vision: a Geometric Viewpoint*. MIT Press, 1993.

[4] O. Faugeras and B. Mourrain. About the correspondences of points between *N* images. In *Proceedings of the IEEE Workshop on Representation of Visual Scenes*, pages 37–44, 1995.

[5] O. Faugeras and L. Robert. What can two images tell us about a third one? *IJCV*, 18(1):5–19, April 1996.

[6] G. H. Golub and C. van Loan. *Matrix Computations*. Johns Hopkins University Press, 3 edition, 1996.

[7] C. Harris and M. Stephens. A combined corner and edge detector. In *4th Alvey Vision Conference*, pages 147–151, 1988.

[8] R. Hartley. Estimation of relative camera positions for uncalibrated cameras. In *Proc. 2nd European Conf. on Computer Vision*, Lecture Notes in Computer Science 588, pages 579–587, 1992.

[9] R. Hartley. Projective reconstruction and invariants from multiple images. *IEEE Trans. Pattern Analysis and Machine Intell.*, 16(10):1036–1041, 1994.

[10] R. Hartley. Lines and points in three views and the trifocal tensor. *Int. Journal of Computer Vision*, 22(2):125–140, 1997.

[11] R. Hartley. Minimising algebraic error in geometric estimation problem. In *Proc. 6th Int. Conf. on Computer Vision*, pages 469–476, 1998.

[12] T. S. Huang and O. Faugeras. Some properties of the E matrix in two-view motion estimation. *IEEE Trans. Pattern Analysis and Machine Intell.*, 11(12):215–244, 1989.

[13] D. G. Luenberger. *Linear and Nonlinear Programming*. Addison-Wesley, USA, second edition, 1984.

[14] S. Maybank and O. Faugeras. A theory of self-calibration of a moving camera. *Int. Journal of Computer Vision*, 8(2):123–151, 1992.

[15] M. Pollefeys, R. Khoch, and L. Van Gool. Self-calibration and metric reconstruction in spite of varying and unkown internal camera parameters. In *Proc. 6th Int. Conf. on Computer Vision*, pages 90–95, 1998.

[16] M. Pollefeys and L. Van Gool. Self-calibration and metric reconstruction in spite of varying and unkown internal camera parameters. Technical Report TR9707, ESAT-MI2, K. U. Leuven, 1997.

[17] P. Sturm. Critical motion sequences for monocular self-calibration and uncalibrated euclidean reconstruction. In *Proc. Conf. Computer Vision and Pattern Recognition*, pages 1100–1105, 1997.

[18] R. Szeliski and P. H. S. Torr. Geometrically constrained structure from motion: Points on planes. In *3D Structure from Multiple Images of Large-Scale Environments, European Workshop SMILE'98*, Lecture Notes in Computer Science 1506, pages 171–186, 1998.

[19] P. H. S. Torr and A. Zisserman. Robust parametrization and computation of the trifocal tensor. *Image and Vision Computing*, 15(8):591–605, 1997.

[20] B. Triggs. Autocalibration and the absolute quadric. In *Proc. Conf. Computer Vision and Pattern Recognition*, pages 609–614, 1997.

[21] R. Tsai. An efficient and accurate camera calibration technique for 3D machine vision. In *Proc. Conf. Computer Vision and Pattern Recognition*, pages 364–374, 1986.

[22] R. Tsai and R. Lenz. Techniques for calibration of the scale factor and image center for high accuracy 3D machine vision metrology. *IEEE Trans. Pattern Analysis and Machine Intell.*, 10(5):713–720, 1988.

[23] A. Zisserman, D. Liebowitz, and M. Armstrong. Resolving ambiguities in auto-calibration. *Phil. Trans. Royal Soc. London A*, 356:1193–1211, 1998.